

Visualization of semantic differential studies with a large number of images, participants and attributes

Akari Iijima
Ochanomizu University
Tokyo, Japan
ijima.akari@is.ocha.ac.jp

Takayuki Itoh
Ochanomizu University
Tokyo, Japan
itot@itolab.is.ocha.ac.jp

Hsiang-Yun Wu
TU Wien
Wien, Austria
hsiang.yun.wu@acm.org

Nicolas Grossmann
TU Wien
Wien, Austria
ngrossmann@cg.tuwien.ac.at

Abstract—The Semantic Differential (SD) Method is a rating scale to measure the semantics. Attributes of SD are constructed by collecting the responses of participant’s impressions of the objects expressed through Likert scales representing multiple contrasting with some adjective pairs, for example, dark and bright, formal and casual, etc. Impression evaluation can be used as an index that reflects a human subjective feelings to some extent. Impression evaluations using the SD method consist of the responses of many participants, and therefore, the individual differences in the impressions of the participants greatly affect the content of the data. In this study, we propose a visualization system to analyze three aspects of SD, objects (images), participants, and attributes defined by adjective pairs. We visualize the impression evaluation data by applying dimension reduction so that, users can discover the trends and outliers of the data, such as images that are hard to judge or participants that act unpredictably. The system firstly visualizes the attributes or color distribution of the images by applying a dimensional reduction method to the impression or RGB values of each image. Then, our approach displays the average and median of each attribute near the images. This way, we can visualize the three aspects of objects, participants and attributes on a single screen and observe the relationships between image features and user impressions / attribute space. We introduce visualization examples of our system with the dataset inviting 21 participants who performed impression evaluations with 300 clothing images.

Index Terms—Semantic Differential Method, Visualization

I. Introduction

The Semantic Differential (SD) Method [Osg52] is an important evaluation methodology in many academic and industrial fields. Academic fields in which SD is applied include psychological experiments and emotional information processing. SD has been also industrially applied for advertising strategies and customer analysis.

Impression evaluation results can be used as training datasets for machine learning systems that estimate the impression of digital content. We expect the impression of the images in the test datasets can be appropriately estimated by machine learning systems by applying the images those impressions are annotated as training datasets. In this case, the quality of impression estimation is greatly affected by the quality of the training datasets. Therefore, users need to understand the distribution of impression evaluation results in the training datasets. The accountability of the behavior of machine learning systems

is often critical. Understanding the distribution of training data is important from the viewpoint of the accountability of machine learning systems.

We have focused on impression evaluation in the form of presenting images to participants and hearing their impressions. Since the responses of impression evaluations vary person-by-person, it is usually required to gather a large number of participants. In addition, a large number of objects, namely a large number of images in this paper, are required to aggregate the impression evaluation results and perform statistical processing. Furthermore, participants who look at images may perceive a wide variety of impressions, and therefore it is necessary to prepare many questions for a single image. In other words, we need to invite many participants and prepare many attributes in an experiment of impression evaluation for images. For this reason, impression evaluation is called multi-dimensional impression evaluation in this paper.

SD is used in a wide range of fields, but the development of visualization methods for SD is still an open problem. Here, we need to focus on three aspects, objects (images), participants and attributes, (in this paper, these are called three aspects.) because one of the mainstreams of the visualization for SD is two-aspect analysis[BL94]. A typical example of two-aspect analysis deals with only participants and attributes from the three aspects. We may need to show multiple analysis results (e.g. spatial diagrams and tables) while focusing on the three aspects by the combination of multiple two-aspect analyses. On the contrary, our goal is to visualize the three aspects on a single screen with a combination of SD and dimension reduction.

Questions for impression evaluations are often prepared in n -point Likert scale in order to collect the evaluation results in tabular forms. We adopt the SD method [Osg52] to the questions of impression evaluation and ask participants to answer the conformity with pre-defined adjective pairs in n -point Likert scale. This study presents multiple attributes (questions in this study) to participants for each of the images and treats the results obtained from these responses as three-dimensional data with many images, many participants, and multiple attributes as three aspects.

Analysis of such impression evaluation results requires a multifaceted viewpoint. Focusing on the images, for example, users can observe what kind of images have similar responses and what kind of images are easy to answer their impressions. Focusing on the participants, for example, they can observe which groups of participants have similar responses and which participants have drastically different responses against other participants. Focusing on the questions, for example, they can observe which attributes have correlations with others and which attributes are easy to answer.

Based on these observations, we propose a visualization system that assists the understanding of the distribution of impression evaluation results. Our study presented in this paper consists of the following three processing steps:

[Step 1] Perform impression evaluations of a set of images applying SD.

[Step 2] Visualize the distribution of impressions or colors by applying a dimensionality reduction method to impression and RGB values of the images.

[Step 3] Observe the association between image features and evaluation values while the presented system displays the average and median of each attribute near each of the images.

Our main contribution is as follows. The proposed visualization system enables data owner to analyze impression evaluation data from the viewpoints of the following three aspects. The system visualizes all three aspects in a single display space.

The remainder of this paper is structured as follows. Section II describes related research, and Section III describes the proposed method. Section IV presents visualization examples and discussion on this study, and finally, Section V summarizes our study and mentions our future work.

II. Related Work

SD is proposed by Osgood [Osg52], an American psychologist, to SD requires participants to answer the degree of fitness to adjective pairs (attributes), for example, dark and bright, formal and casual, etc. Analysts use SD to determine the attributes from the average value and median values of responses of the participants. Likert scale is often applied as the responses form and we assume that the level of the interval scale is satisfied.

There have been several studies [DWS16], [SDH14] on analyzing impression evaluation data applying the SD method as follows. One of the open problems in these studies is to determine attributes only with real values of the Likert scale used to measure the attributes of SD. To solve this problem, Matsuo et al. [Mat+10] presented a method to analyze SD using a Fuzzy decision tree. This method extracts the partial evaluation attributes that affect the overall evaluation from the fuzzy decision tree.

Three aspect factor analysis [Dai82] is a data analysis method that takes into account individual differences in the responses of the SD. Several other studies also focused on the scales of these three aspects, objects, participants. For example, Toyoda et al. [Toy01] presented an exploratory positioning analysis for the three aspects. However, these studies require two views, a table of factor loading of a certain aspect and the two-dimensional space (factor space [BL94]) generated by a dimension reduction method [Kan+18]. Based on this discussion, we propose a visualization system with a single display space that combines a view for objects (images) and views for an exploratory positioning analysis.

Factorization is also a useful tool for impression evaluation. Stoklasa et al. [STS19] applied the factorization focusing only on attributes. Meanwhile, Osgood applied it to the analysis that breaks down the attribute into three basic factors (evaluation factor, potency factor, and activity factor) [Osg64]. Our study has not yet applied the factor decomposition for attributes, but it may be effective to apply as future work.

We selected Principal Component Analysis (PCA) from many existing dimension reduction subjectively. Our visualization system maps a lot of images onto the display space based on the dimension reduction results, and therefore we preferred PCA because it brings well-spread results. Dimension reduction methods such as PCA have been applied for browsing a large number of images. Here, user interaction is one of the most important aspects [Mog+] while observing the distribution of the displayed images. For example, we may need to click particular images and show the detailed information of the specified images. Image browsing applying the dimension reduction methods have problems while such user interactions with the images. Our implementation reduces the overlaps among images by using a visualization library supporting rich interactive operation functions.

III. Visualization of Impression Evaluation

This section presents our study on the visualization of multi-dimensional impression evaluation data. This study supposes that many participants evaluate a set of images by applying SD. Our system visualizes impression and color distributions of the set of images by applying a dimension reduction method to the impression and RGB values of the images. Here, users can observe the association between image features and evaluations using the visualization system by displaying the average and median of the attributes near the images.

This section presents the processing flow of the proposed system. Section III-A describes the format of impression evaluation datasets, and then, Section III-B presents the dimension reduction process for impression and RGB values of the images. Section III-C illustrates how to display the average and median values of the attributes,

and finally, Section III-D presents the visualization of the characteristics of personal evaluations.

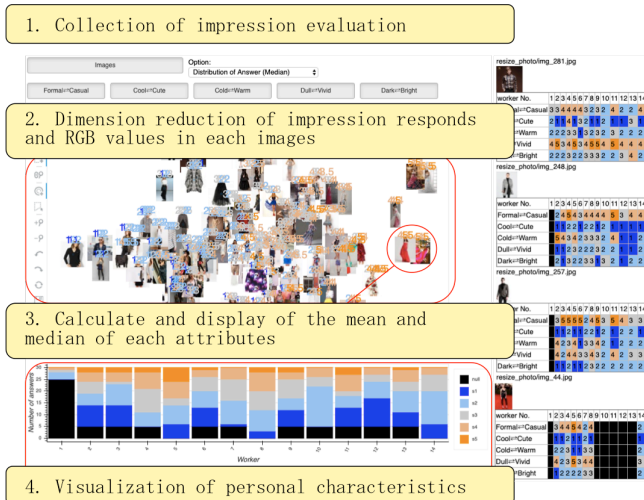


Fig. 1. Proprietary implemented visualization system.

A. Collection of Impression Evaluation and Construction of Data

In this study, it is required to prepare images and set multiple adjective pairs appropriate for the content of the images as attributes. Then, the study requires multiple participants to answer their relevance to the adjective pairs to each of the images. From the set of responses collected as described above, the study supposes to construct impression evaluation datasets with the three aspects. We applied the five-point Likert scale as the response format and calculated the average and median values for each attribute of each image from the responses of participants in this study.

B. Dimension Reduction of Impression and RGB Values

Next, the system sets multidimensional vectors consisting of the average and median of the attributes to a set of images and applies Principal Component Analysis (PCA) to the vectors to display the image into a two-dimensional display space. As a result, similarly impressed images are placed closer together on the display space. The system also applies PCA to RGB values of the images so that similarly colored images are placed closer to each other. Users can observe two types of image distribution and find the similarity of impressions or colors. They can select image placement based on the following three values, "impression value (average value)", "impression value (median value)", and "RGB value", by pressing the button at the upper left of the visualization system as shown in Fig. 1.

C. Display of the Mean and Median of Each Attributes

The system displays the average and median values of an attribute are near the images after placing in the display space applying PCA. Individual colors with two hues (blue

and orange in our implementation) are assigned to the points (1 to 5) on the Likert scale as shown in Fig. 2. Here, the average and median values are rounded off to integer values. Fig. 3 shows an example that the median values of an attribute are displayed in order.



Fig. 2. Five-scale color setting.



Fig. 3. The median or average value is displayed in the order of each attribute near the image.

D. Visualization of Personal Characteristics

In order to understand the personal characteristics of participants' responses, the system visualize the details of the responses in the five-point Likert scale using a stacked bar graph of the total number of responses. An example is shown in the lower-left part of Fig. 1. The horizontal axis shows the participants in the order of their sequential identifiers, while the vertical axis shows the total number of responses. Unanswered responses are drawn with black color, and responses (1 to 5) are painted in the colors specified in Fig. 2. In addition, when a user selects an image for which user want to know the details, the system displays the user-selected image and its details of the responses as shown on the right of Fig. 1.

IV. Example of Visualization by Using Our Technique

We implemented the presented visualization system by extending the Python visualization library "Bokeh". Bokeh is a software library that supports rich interactive operations such as viewpoint operations (translation and scaling), object selection by clicking operations, and saving the visualization results as images. We randomly collected clothing images from Deep Fashion Database ¹ to generate this visualization example. We also prepared two types of datasets as impression evaluation datasets as shown in Table 1. In dataset 1, fourteen Japanese women, aged 20 to 25 years, participated in this experiment. In dataset 2, twenty-one people, aged 21 to 57 years (6 men and 15 women, 13 Japanese and 8 non-Japanese), participated. Especially when it comes to non-Japanese, they are living in Austria and their nationalities vary (except Japan).

¹<http://mmlab.ie.cuhk.edu.hk/projects/DeepFashion.html>

The selection of the attributes is based on Furukawa et al. [Fur+17]. We chose five adjective pairs based on our subjective assumptions that these adjective pairs have low correlations.

In this example, we first visualized the dataset 1 and analyzed the results, and then, modified some of the attributes of the dataset 1 based on the analysis results, and created the dataset 2 while increasing the number of images and the number and variety of participants. We can observe the impression evaluation results from the following three viewpoints. The difference between the two datasets will be discussed Section IV-B.

- Unique clusters of images that have similar responses or RGB values
- Appropriate attributes
- Characteristics of responses of participants

Section A shows the visualization results and discusses the impression evaluation of the dataset 1. Section B discusses the visualization results of the dataset 2. Since the number of images, attributes and types of participants are different between the two data, we will observe them separately.

TABLE I
IMPRESSION EVALUATION DATA

	Dataset 1	Dataset 2
Image	137	300
Participant	14 women (Japanese)	15 women, 6 men (13 Japanese, 8 non-Japanese)
Attribute 1	Formal - Casual	Formal - Casual
Attribute 2	Cool - Cute	Masculine - Feminine
Attribute 3	Cold - Warm	Modern - Classic
Attribute 4	Dull - Vivid	Dull - Vivid
Attribute 5	Dark - Bright	Simple - Gorgeous
Likert scale	5	5

A. Visualization of the Dataset 1

Figs. 4 to 7 show the results of visualizing impression evaluations stored in the dataset 1. Fig. 4 shows the overall visualization results where similarly impressed images are placed nearby. Real values colored as explained in Fig. 2 indicate the average values. Blue letters occupy a large part of the visualization result that illustrates many participants answered as 1 or 2.

Fig. 5 shows a visualization example that displays only the median of attribute one by one without the images. We can observe that similarly impressed images are closely placed in the display space from this visualization example. Note that Fig. 5 (upper) shows the visualization with the attribute "Formal-Casual," while Fig. 5 (lower) shows the visualization with the attribute "Cold-Warm." We can see observe that many participants answered as 1 or 5 for Formal-Casual, whereas they answered as 2, 3, or 4 for Cold-Warm. The results demonstrate that the former



Fig. 4. Visualization result displaying the average value near the image.

adjective pair is easier to get clear impressions rather than the latter one.

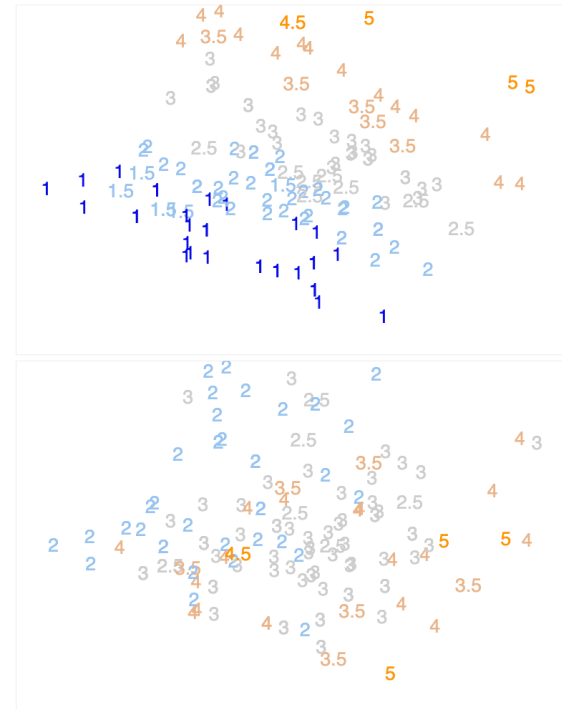


Fig. 5. (Upper)Attribute Formal-Casual (Lower)Attribute Cold-Warm median of each image.

Our implementation displays the detailed of the responses as shown in Fig. 6, when a user wants to see such detailed information for a particular image shown in Fig. 4 and actually selects the image. This visualization result shows the two images in Fig. 6 (left) were rated as 4 or 5 for the attribute "Dark-Bright." This means many participants had a bright impression of both images. Even though these images had different colors of clothes and backgrounds, participants had a bright impression of both images. This suggests that interpretations of the attribute "Dark-Bright" may differ among the participants.

Fig. 7 shows an example that visualizes the characteristics of participants' responses in the five-point Likert

scale. The color map shown in Fig. 2 is applied for the five-point evaluation, where unanswered attributes are painted in black. Participants are arranged along the horizontal axis in Fig. 7 in the order of their sequential identifiers, while the vertical axis represents the total number of responses. Here, it is clearly visualized that participants 5 and 7 often answer as 3 which is the middle of the evaluation scale, indicating that they had an aggressive impression evaluation. Conversely, participants 11 and 13 often answer as 1 and 5, indicating that they had a conservative impression evaluation. Participant 1 has many unanswered questions, and therefore we needed to promptly perform the impression evaluation.

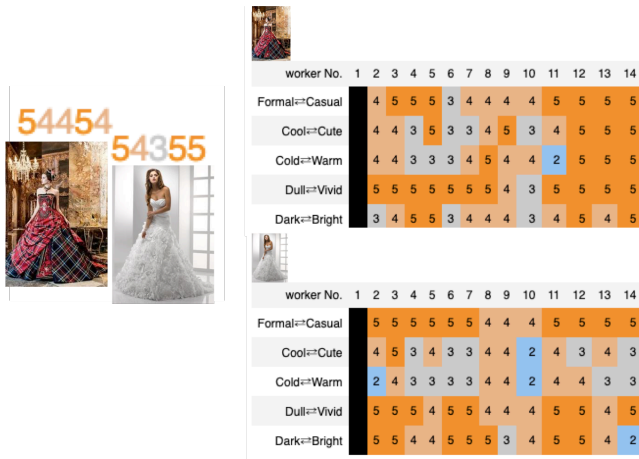


Fig. 6. (Left) Selected images (Right) Details of the response result for the selected image.

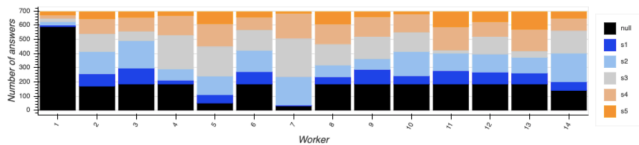


Fig. 7. (Dataset 1) Trends in participants' responses on a five-scale.

B. Visualization of the Dataset 2

We observed the following issues through the visualization of the dataset 1:

- Attributes needs to be reselected. We found that participants might associate multiple meanings with several adjective pairs through the visualization of the dataset 1. We suppose participants might have two possible interpretations for the adjective pair Cold-Warm: whether the materials of clothes seem cold or warm, and whether the atmosphere of clothes looks cold or warm. We also assumed participants might associate two meanings with the adjective pair Dark-Bright.
- Since the participants were only Japanese female students in their twenties, it is not sufficient to generalize the discussion solely from this dataset.

From the above viewpoints, we created the dataset 2 that increased the number of images and updated the problematic attributes. In addition, the number of male participants in the dataset was increased as well as female, and the number of participants from overseas was also increased as well as Japanese. Note that at the time of submission of this manuscript, responses are still being collected, and therefore, many unanswered questions remain in the dataset.

Figs. 8 to 13 show the visualizing results of the dataset 2. Here, similarly colored images are placed close in the visualization results shown in Figs. 8 and 9. We can observe results with different responses for similar images in Modern-Classic attribute. Fig. 9 (upper) shows a visualization result that displays the median of the responses for the attribute Modern-Classic. Fig. 9 (lower) shows an enlarged view of the image surrounded by red circles and the median of the responses. Regarding the two clothing images on the left, both tops and bottoms have the same colors and shapes, suggesting that the responses of the impression for both images were similar. On the other hand, the median of the responses had a large difference between the images on the right, even though both two clothes images take white dresses. This suggests that the impression values may greatly differ even if the clothes taken in the images have the same colors and slightly different shapes. Here, we remarked that different persons are invited as models in the clothing image data used in this study. Hence, it suggests the possibility that faces, expressions, hairstyles, and skin colors of the models might affect the impressions of the images.



Fig. 8. Image distribution where images with similar RGB values are placed nearby.

Fig. 10 shows an example that visualizes the variation of participants' responses in the five-point evaluation. The colormap is as described in Section III-C. The dataset 2 contains a smaller number of responses for each image since there are many black regions in Fig. 10. Therefore, from the current visualization results, individual responses may have greater efforts on the visualization results than the dataset 1.

In the following, similarly impressed images are placed close in the display space in Figs. 11 to 13. Figs. 11



Fig. 9. (Upper) The median value of attribute Modern-Classic for each image. (Lower) Enlarged image surrounded by red circles, and each median.

and 12 show the visualization results that represent the differences of responses between Japanese and non-Japanese participants. From these visualization results, we could clearly observe the differences of distributions of responses, especially with the attributes Dull-Vivid and Modern-Classic. We can observe that Japanese participants are relatively likely to answer as 2, 3, and 4 in the five-point, while overseas participants are most likely to answer as 1 or 5. It means that Japanese participants often had modest responses compared to the participants in other countries. In addition, it is possible that the interpretation of impressions between overseas and Japan is particularly different between the attributes Dull-Vivid and Modern-Classic. We suppose it is sometimes difficult for Japanese participants to answer clear impressions for these attributes compared to overseas participants.

Fig. 13 shows the visualization results that represent the difference of responses between male and female participants. Fig. 13 (upper) shows the median value of responses of female participants for the attribute Formal-Casual, while Fig. 13 (lower) shows the median of the responses of male participants. The median values are displayed in blue as they are 1 or 2 corresponding to Formal, and conversely, they are displayed in orange as they are 4 or 5 corresponding to Casual. The images surrounded by red circles in Fig. IV-B are the same image in the responses of both male and female participants. Despite the same image, many female participants answered as Casual, while many male participants answered as Formal. From this result, we found the attribute Formal-

Casual associates different interpretations between male and female participants. However, as discussed with Fig. 10, a smaller number of participants had responses to each image. User can discover which participants are answering and how much. The the impression value of each individual might greatly reflect in the visualization results. Alternatively, it is possible that the participants gave incorrect responses to the adjective pairs.

The system visualizes all three aspects mentioned in Section 1: objects (images), participants, and attributes in a single display space. In addition, data owners can analyze impression evaluation data from the viewpoints of three aspects. In particular, attribute analysis is one of the strengths of the proposed system. Users can discover and analyze whether the interpretation of impression items differs between participants. This may be particularly suitable for performing pilot tests on small amounts of data before performing large-scale experiments.

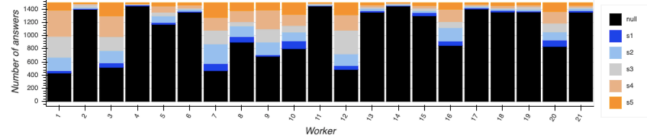


Fig. 10. (Dataset 2) Trends in participants' responses on a five-scale.

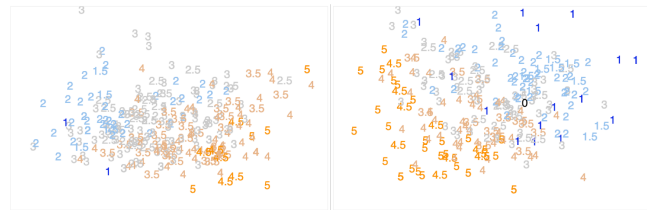


Fig. 11. Median for each image in attribute Dull-Vivid (Left) Japanese responses and (Right) Overseas responses.

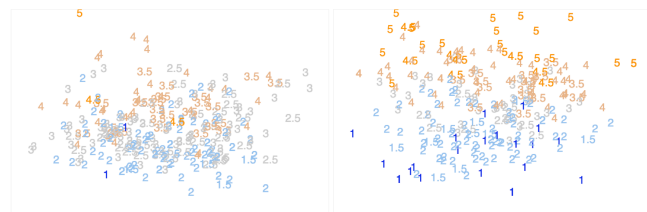


Fig. 12. Median for each image in attribute Modern-Classic (Left) Japanese responses and (Right) Overseas responses.

V. Conclusion and Future Work

This paper presented our study on multi-dimensional visualization of impression evaluation. In this study, we performed an impression evaluation using SD with a large number of images and then visualized the distribution of impressions and colors of the images by applying PCA. The visualization system implemented for this study displays the average and median of each attribute near images on the two-dimensional space so that associations



Fig. 13. Median for each image in attribute Formal-Casual (Upper)women’s responses and (Lower)men’s responses.

between image features and participants’ responses can be observed. In addition, the system displays the details of responses and the distributions of participants’ responses, when a user selects an image for which he/she wants to know the details.

Our future issues or improvement are as follows.

- Creation of larger datasets of impression evaluation for images
- Development of more scalable visualization methods
- Definition of the annotation format as training data for machine learning

A larger amount of images would be required so that the datasets can be used as training data for machine learning. Our final goal is to develop a database of fashion preferences that can be converted into datasets for machine learning systems. It is not a straightforward issue since images we used are different in terms of colors and designs in fashion, the ethnicity of the models, the poses of the fashion models, and the skill of the photographers. We need to further discuss how to develop the database. In addition, participants need to be unbiased, and various types of participants should submit an equal number of responses, in order to improve the reliability of the data. We plan to create larger datasets and proceed larger-scale impression evaluations with the above points in mind.

Second, the current visualization is not sufficiently scalable because the data form is fixed. Therefore, it is necessary to extend the visualization method. Finally, we will define the format of the impression evaluation results and add to the images as annotations. This is necessary to use image impression evaluation results as training data for machine learning.

Acknowledgment

We appreciate ABEJA, Inc. for introducing the image dataset. We also appreciate Dr. Manuela Waldner, TU Wien, for overall advises and suggestions on implementation and visualization design.

References

- [Osg52] Charles E Osgood. “The nature and measurement of meaning.” In: *Psychological bulletin* 49.3 (1952), p. 197.
- [Osg64] Charles E Osgood. “Semantic differential technique in the comparative study of cultures”. In: *American Anthropologist* 66.3 (1964), pp. 171–200.
- [Dai82] Kiichi Dai. “Application of the three-mode factor analysis to industrial design of chair styles.” In: *Japanese Psychological Review* (1982).
- [BL94] Margaret M Bradley and Peter J Lang. “Measuring emotion: the self-assessment manikin and the semantic differential”. In: *Journal of behavior therapy and experimental psychiatry* 25.1 (1994), pp. 49–59.
- [Toy01] H Toyoda. “An exploratory positioning analysis: Three-mode multivariate analysis for semantic differential data”. In: *Shinrigaku kenkyu: The Japanese journal of psychology* 72.3 (2001), pp. 213–218.
- [Mat+10] Takeshi Matsuo et al. “Sensitivity Information Analysis of Running Shoes Using Fuzzy Decision Tree and Visualization of Analytical Results”. In: *SCIS & ISIS SCIS & ISIS 2010. Japan Society for Fuzzy Theory and Intelligent Informatics. 2010*, pp. 156–160.
- [SDH14] Hidenori Sakanwia, Fangyan Dong, and Kaoru Hirota. “A Visualization Method of Kansei Texture and Its Individual Difference.” In: *Journal of Information Assurance & Security* 9.7 (2014).
- [DWS16] Ryuichi Doizaki, Junji Watanabe, and Maki Sakamoto. “Automatic estimation of multidimensional ratings from a single sound-symbolic word and word-based visualization of tactile perceptual space”. In: *IEEE transactions on haptics* 10.2 (2016), pp. 173–182.

- [Fur+17] Takao Furukawa et al. “Quantitative trend analysis of luxury fashion based on visual impressions of young Japanese women”. In: *International Journal of Fashion Design, Technology and Education* 10.2 (2017), pp. 146–157.
- [Kan+18] Yoichiro Kanno et al. “Visualization of Flavor of Sake by Sensory Evaluation and Statistical Method”. In: *Sensors and Materials* 30.5 (2018), pp. 1197–1206.
- [STS19] Jan Stoklasa, Tomáš Talášek, and Jana Stoklasová. “Semantic differential for the twenty-first century: scale relevance and uncertainty entering the semantic space”. In: *Quality & Quantity* 53.1 (2019), pp. 435–448.
- [Mog+] Baback Moghaddam et al. “Visualization & Layout for Image Libraries”. In: ().